DOCUMENT RESUME

ED 102 201 TM 004 258

AUTHOR Simon, Alan J.; Joiner, Lee M.
TITLE Adapting the Peabody Picture Vocabulary Test for Use With Mexican Children.
PUB DATE [Apr 74]
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Culture Free Tests; Elementary Education; Elementary School Students; *Intelligence Tests; Item Analysis; *Mexicans; Statistical Analysis; *Test Bias; *Test Reliability; Test Validity; *Translation
IDENTIFIERS *Peabody Picture Vocabulary Test

ABSTRACT

The effectiveness of test adaptation based on item selection and reordering of a Spanish (Mexican) version of the Peabody Picture Vocabulary Test (PPVT) was examined. Translated forms were administered to a sample of Mexican students. One item from each pair (A and B) was selected and reordered using a priori rules. The revised instrument was administered to a new cross-validation sample. Findings confirmed the cost effectiveness of this technique for improving reliability and validity over simple translation or the creation of completely new items for populations of different culture and language. (Author)

ED102201

Adapting The Peabody Picture Vocabulary Test For Use With Mexican Children

Alan J. Simon, Ph.D.
Queens College, C.U.N.Y.

Lee M. Joiner, Ph.D.
Southern Illinois University

TM004 258

A Paper Presented at The Annual Convention
of The American Educational Research Association
Chicago, Illinois April 1974

Recent litigation involving children who were identified as mentally retarded by school systems has posed serious problems for educators and psychologists who routinely use tests for determining educational placement. A prime grievance of minority group parents has been that commonly used intelligence tests are unfair because of language and cultural biases. In the District Court of Northern California, a suit was filed on behalf of nine Mexican-American students whose primary language was Spanish but who were tested in English, and whose scores were sufficiently low (IQ 30-72) to warrant placement in classes for the mentally retarded. When retesting was conducted in Spanish, their average score increased 15 points and seven of the nine were above the mental retardation cutting point. On February 5, 1970 a stipulated agreement order was consummated which, among other things, required that children be tested in their primary language. Bilingual examiners and interpreters were to be used to implement this agreement (Weintraub and Abeson, 1974).

The problem that is encountered in attempting to implement decisions that mandate fairer testing procedures for Mexican-American children is that there is a scarcity of tests that are comparable across languages and cultures. Simple translations of achievement or intelligence tests result in tests having lower reliability, validity, means, and standard deviations.

Yet an examination of a collection of currently available tests (ETS, Test Collection, 1972) reveals that in almost every case, researchers have chosen to modify existing tests rather than become involved in the

costly highly technical and very time consuming task of developing new tests. The majority of these modifications take the form of simple translations. The Psychological Corporation catalogue is offered as but one example. It identifies nine tests on its list of "Spanish-Language Editions of Tests". Of these, seven were simply translated, and two, the WISC and the WAIS were "adapted to the Spanish Culture". In the U. S., researchers have often simply translated tests for use with non-English speaking children (Diether, 1968).

After studying intelligence tests translated from English to Spanish as spoken in Puerto Rico, Rosa (1955) concluded that one of the main problems in translating tests was that a translated word, although expressing identical concepts, may be of a different degree of difficulty in the new than in the old language. Symeo (1970) noted the problem that a concept or word may not be present in the new culture; the translation, then is accomplished by replacing the intended concept with one which is judged to be similar, this judgement being subject to debate. Another problem is that a word may possess a single meaning in one culture but possess multiple meanings in the other; or a word may have opposite meanings.

If adequate translation of tests is difficult to accomplish, neither will more extensive adaptations be free of problems. In this respect the importance of Ali's (1967) findings concerning the use of an American achievement test in East Pakistan can be seen. Item analysis showed that the translated version of the tests made a greater contribution to the research than did the adapted version. Adaptation

of test items in terms of item, word, and structural analysis did seem worth pursuing as a means of developing tests until and unless the criteria of and the elements for adaptation were determined through further experimentation. Attempting adaptation along with translation of a standardized multicultural achievement test on a priori items without empirical evidence was not "cost-efficient" in improving test reliability or validity. This does not mean that straight translation was a satisfactory solution, however.

In an attempt to determine the statistical equivalence of test items in a Portuguese translation of a Spanish test, Clark (1969) made the assumption that equivalence could be demonstrated if corresponding items were symmetric with regard to certain characteristics. Among the characteristics considered relevant were: D-BIP reliability; item difficulty; and item x total test correlations. On the basis of statistical consistencies, Clark concluded that the common linguistic principles of the two languages permitted the successful item translation/adaptability. He suggested that with less related languages the use of direct translation would be a more difficult undertaking.

Rose (1969) and Benoit and Fuchs (1969) referred to the possible need for reordering of test items in a homogenous test based on an analysis of the correlation between item difficulty and item validity. It is the contention of the present authors that if a wide range test such as the PPVT is to be useful, then the basal and ceiling point rationale must be operative in the foreign language version. This type of structure presupposes a near perfect correlation between item order item difficulty.

The major purpose of the present study was to determine whether an improved Mexican version of the Wechsler Preschool and Primary Scale of Intelligence could be constructed by directly translating test items 1 and 2 of the American test and then using a set of decision procedures to select the better item from the item pairs. A by-product of the selected items would then be a comparison of the psychometric properties of the two versions of the WPPSI. The matching pairs Mexican version would be administered to Mexican samples to determine whether or not its psychometric properties were significantly superior to those of the simple translations. A second concern of the study was to note the extent of changes in the psychometric properties of the test when only simple translations were used.

Method

Translating the WPPSI

Items 1 and 2 of the WPPSI were translated by three bilingual native Mexicans who were members of the faculty of the University of New Mexico, Mexico. Each proceeded independently, but in the final stage worked as a group to achieve consensus. These translated items were referred to as Item S_1 and S_2.

Item Pairs Selected

Thirty-three students in each of grades one through six were randomly selected from the rural school population of Oaxaca, Mexico, a metropolitan center of 210,000 inhabitants. Representativeness in each school, and within an otherwise arbitrary selection was achieved through

[illegible] Investigations were delineated by graduate students in the Department of Sociology of the University of Iowa [illegible]. Subjects responded to both items I_1 and I_2 with a percent of the sample objecting between items. One half of the is responded to item I_1 first and the other half to I_2 to preclude learning effect systematically biasing the results for one of the items.

Constructing the Guttman Scale, i.e. $(X)(Y)$

Using the ordered criteria listed below, one item was selected from the pool (items I_1 and I_2) for inclusion in the Guttman instrument (item I_1):

1. Based on total consonance between item score and total test score. This characteristic contributes to the factor structure of the items as well as to item reliability and was reported as the single most important characteristic of the items. (Kennedy, 1972).
2. When an decision could be made using criterion #1, the point-biserial correlation between item score and chronological age for all students in the sample was used as a second assessment standard used in an attempt to increase such reliability.
3. If still no choice was possible, item difficulty as indicated by the mean for each item was examined. In order to maximize reliability, the item of each pair whose percent score was closest to .50 was chosen.

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		

REFERENCES

Ali, M.B. To establish the feasibility of using translated and adapted versions of an American-made mathematics achievement test in East Pakistan. Dissertation Abstracts, 1967, 28 (1-A), 115-116.

Clark, John. Item by item parallel achievement test for different foreign languages. Journal of Educational Measurement, Volume 2, Number 1, 1964, 77-83

Dunn, L.M., Manual for Peabody Picture Vocabulary Test, American Guidance Service, Minnesota, 1965.

Educational Testing Service. "Test Collection", Princeton, N.J., (1972).

Helmstadter, G.C. Principles of Psychological Measurement. New York, Appleton-Century-Crofts. 1964.

Hyness, D. Linguistic aspects of comparative political research, in Holt, R. T. and Turner, J. E. (ed.), op. cit.

McNemar, Q., The Revision of the Stanford-Binet Scale, an Analysis of the Standardization Data. Boston, Houghton Mifflin, 1942.

Nunnally, J.C. Psychometric Theory. New York, McGraw-Hill, 1967.

Psychological Corporation. Test Catalogue. New York, 1972.

Renzulli, J. S., Paulus, D. H. A cross-validation study of the Item Ordering of the Peabody Picture Vocabulary Test. Journal of Educational Measurement, 1969, 6 (1), 15-20.

Rieber, M. and Womack, M. The Intelligence of Preschool Children as Related to Ethnic and Demographic Variables. Exceptional Children, 1968, 34, 8, 609-614.

Roca, P. Problems of adapting intelligence scales from one culture to another. Oficina de Evaluation, Departmento de Instruccion Publica, Halta Rey, Puerto Rico, 1960.